The genomic landscape of head and neck squamous cell cancer and their implication for treatment in NGS-based umbrella trial

Shinwon Hwang

Department of Medicine

The Graduate School, Yonsei University

The genomic landscape of head and neck squamous cell cancer and their implication for treatment in NGS-based umbrella trial

Directed by Professor Sangwoo Kim

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Medical Science

Shinwon Hwang

December 2023

This certifies that the Doctoral Dissertation of
Shinwon Hwang is approved.

Thesis Supervisor : Sangwoo Kim

Thesis Committee Member #1: Hye Ryun Kim

Thesis Committee Member #2: Tae-Min Kim

Thesis Committee Member #3: Nam Suk Sim

Thesis Committee Member #4: Dongwan Hong

The Graduate School
Yonsei University

December 2023

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to Professor Sangwoo Kim. Working under his guidance in TGIL has been an honor and has significantly contributed to my growth as a scientist.

I am also immensely grateful to Professor Sang Ho Oh for his continuous guidance and invaluable advice throughout my research journey. Additionally, I appreciate Tae-Gyun Kim for his advice on research and for recommending the TGIL lab.

My sincere thanks go to Professor Hye Ryun Kim for providing me with the opportunity to embark on this work. I am thankful to the KCSG groups for allowing me to conduct this research under their expansive umbrella trial. Furthermore, I express my gratitude to Hye Ryun Kim, Tae-Min Kim, Nam Suk Sim, and Dongwan Hong for their role in advancing my research as thesis examiners.

I would like to express my appreciation to all TGIL lab members for their advice and support. Special thanks to In Seok Yang for his substantial support in lab work and his sincere attitude; Se-Young Jo for his enriching discussions, advice on research and visualization, and for the opportunity to experience F1; Yoo Jin Ha for her guidance in research and support in my endeavors; Eunyoung Kim for her leadership in the lab and insightful advice; Jiyun Hong for his assistance in various aspects of my research; Jeongsoo Won for his help during his tenure as lab head and for his considerate presence in the lab; Seunghee Moon for her assistance in lab work and meticulous attention to detail; Ah Yeon Kim for making laboratory activities enjoyable, especially as a neighbor sitting nearby; Jiho Park for her

bright mood in TGIL and her extensive support during my time as lab leader; Seungseok Kang for his support and collaboration in the lab; Seungho Oh for managing the lab server throughout my time at TGIL; Young-soo Chung demonstrating a lot to learn from planning and bringing in valuable knowledge; Dongsun Kim for his assistance in my studies and for being a great betting partner; Joonho Kim for engaging discussions on sports and neurology; Sangbo Lee for his knowledges of mathematics and future pediatric consultations; Dongwoo Park for his careful help and role as a lubricant in the TGIL lab; Yongjun Kim for his help in this work and the linked-read project as one team; Eunwoo Choi for her help in the lab and for envisioning a bright future for TGIL; Hyeonju Son for aiding my initiation into bioinformatics; Jongwook Oh for his abundant discussions and kind consideration; and Chan Wook Park for our often enlightening coffee time discussions.

My family and friends have been a pillar of support, and I owe them immense gratitude, especially to my father, mother, and sister's family. Lastly, a heartfelt thank you to my fiancé, Jaimie, for her unwavering support and love.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The genomic landscape of head and neck squamous cell cancer and their implication for treatment in NGS-based umbrella trial

Shinwon Hwang

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Sangwoo Kim)

Head and neck squamous carcinoma (HNSCC) originate in the oral cavity, pharynx, and larynx. Recurrent or metastatic HNSCC is associated with a poor prognosis, and treatment options are limited, underscoring the need for enhanced therapeutic strategies. In this study, we analyzed 419 patients with HNSCC enrolled into the TRIUMPH study. Comprehensive analyses, including survival analysis, were conducted to assess the overall genetic landscape, mutational signature patterns, copy number variations (CNVs), and their correlation with patient outcomes. A subgroup analysis was conducted specifically for patients with Human papillomavirus (HPV)-associated oropharyngeal cancer and oral cavity cancer. Aligning with previous studies, our findings confirm genetic aberrations in *TP53, CDKN2A, PIK3CA, FAT1*, and *EGFR*. We identified multiple prognostic factors in patients undergoing targeted therapy. Mutations in the AKT/mTOR pathway were associated with improved survival. Conversely, *NOTCH1* mutations and *MYC* amplification correlated with poorer prognosis in patients receiving PIK3CA-targeted therapies. Additionally, the prognosis of patients treated with CDK 4/6 inhibitors appeared to be influenced by the type of *CDKN2A* gene mutation. Subgroup analysis revealed variations in the genomic landscape based on primary tumor sites, focusing on the relationship between CNVs, *TP53* mutations, smoking status, and age. In subgroup analyses, we focused on the differences in the genomic landscape based on primary tumor sites. In HPV-related oropharyngeal cancer, a distinct exclusivity pattern between *TP53* and *PIK3CA* single nucleotide variants or insertions/deletions was observed. HPV infection status was associated with a favorable prognosis in patients

treated with immunotherapy in the TRIUMPH trial. In oral cavity cancer, the presence of smoking-induced mutations was less pronounced than in laryngeal carcinomas. Also, most young patients with oral cavity cancer had *TP53* mutations, without significant germline variants, smoking history, or HPV infection. These results highlight the clinical importance of detailed genomic profiling in HNSCC, especially in the context of recurrent or metastatic disease, and pinpoint potential targets for personalized therapy.

Key words: head and neck squamous cell cancer, genomic landscape

The genomic landscape of head and neck squamous cell cancer and their implication for treatment in NGS-based umbrella trial

Shinwon Hwang

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Sangwoo Kim)

I. INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) arise in the oral cavity, pharynx, and larynx. HNSCC is the sixth most common cancer in the world and it is influential to account for 5.7% of cancer mortality worldwide [1]. Despite advances in treatment agents such as chemotherapy, immune checkpoint inhibitor, and targeted therapy, patients with recurrent/metastatic HNSCC (R/M HNSCC) have poor survival outcomes, with a median overall survival of around 1 year [2]. The combination of pembrolizumab and chemotherapy or pembrolizumab monotherapy are primarily considered as first-line therapy [3-5]. The cetuximab combined with chemotherapy is considered as an alternative treatment option for patients with R/M HNSCC when immunotherapy is not suitable [6-8]. So far, cetuximab is the only approved targeted therapy for R/M HNSCC patients, but even this has not been treated based on biomarker.

The management of R/M HNSCC is challenging due to the aggressive nature of the disease and limited treatment options. The comprehensive genetic profiling of HNSCCs has been performed through several studies including The Cancer Genome Atlas Network (TCGA) project [9-13]. The key mutations identified in the TCGA dataset include alterations in *TP53, CDKN2A, PIK3CA, NOTCH1, HRAS*, and *CASP8* genes. *TP53* is the most frequently mutated gene in HNSCC, followed by *CDKN2A*, which is

associated with a poorer prognosis. Alterations in *PIK3CA* and *NOTCH1* genes are also common, and their presence has been linked to worse survival outcomes. *HRAS* mutations are more frequently found in HPV-negative HNSCC, while *CASP8* mutations are more frequent in HPV-positive HNSCC. Our previous study on HNSCC using next-generation sequencing (NGS) and RNA expression assay showed that *TP53, CDKN2A, CCND1*, and *PIK3CA* were the most common mutated genes [14]. We also observed distinct gene expression patterns between HPV-positive and HPV-negative HNSCC patients [14]. Despite the comprehensive genetic profiling studies on HNSCC, it remains unclear whether this genetic information can differentially and effectively guide treatment decisions. There is still a need for further research to determine the clinical significance and therapeutic implications of genomic profiling in HNSCC.

We conducted the TRIUMPH (Translational biomarker-driven umbrella project for head and neck and esophageal squamous cell carcinoma) umbrella trial consisting of five therapeutic arms by the Korean Cancer Study Group (KCSG) (NCT03292250) (**Fig. 1).** This study was designed with a scheme allocating four targeted therapies based on relevant pathway mutations according to the NGS results, and immunotherapy was treated when no targetable mutation was found [15]. In this umbrella trial, we screened a total of 419 HNSCC patients, which are the largest numbers reported to date.

Herein, we aim to comprehensively analyze the association between various clinical characteristics and genetic alteration based on NGS data of the largest HNSCC cohort to date. Additionally, we investigated the correlation between survival outcomes and molecular pathway alterations, with the goal of reevaluating their therapeutic value. We aim to identify a potential subpopulation that could be treated with targeted therapy in a cohort combining NGS and clinical data of HNSCC patients through this study.

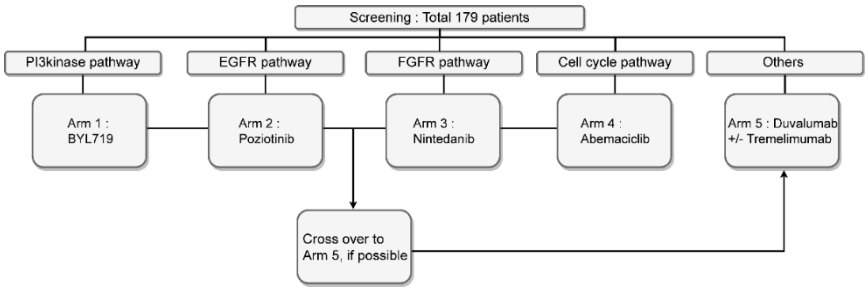

Figure 1. Overall pipeline of TRIUMPH study.

II. MATERIALS AND METHODS

1. Sample and clinical data collection

We collected samples from patients who provided their consent to participate in the TRIUMPH trial screening. The study was authorized by the Institutional Review Board of 19 Korean institutions. The collection of tumor tissue and matched peripheral blood as well as the collection of various clinicopathologic data, such as age, gender, tumor location, use of cigarettes and alcohol, clinical stage, treatment history, and survival data, were all done.

2. Targeted Sequencing, RNA extraction, Immunohistochemistry

Using the Qiagen QIAamp DNA FFPE Tissue Kit, genomic DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) samples in order to sequence 244 genes associated with HNSCC as described in our earlier work (S2 Table) [14]. Using the Agilent SureSelectXT Target Enrichment library production kit, the genomic regions of these genes were isolated, and they were subsequently sequenced on the Illumina HiSeq 2500 platform with a minimum depth of coverage of 1,000X. Immunohistochemistry was used to determine the presence of HPV infection in the samples utilizing p16 expression in the tumor cells.

3. Bioinformatics analysis

A. Data preprocessing and somatic and germline variant calling

The quality of the FASTQ files was assessed based on base quality, GC content, and total base throughput. Trimming was carried out using fastq with several standards, including poly G, length, complexity, and front tail [16]. The Genome Analysis ToolKit (GATK) Best Practices methodology was applied to find somatic and germline variations. In the Genome Analysis Toolkit (GATK) v4.2.3.0, HaplotypeCaller was used to call germline variations, while Mutect2 was used to call somatic variants with the default settings [17]. Hardfilter was used to perform variant filtration, and germline variations that weren't called on typical samples were removed. Minor allele frequencies higher than 0.001 were also used to remove out variations [18]. The annotation of variants was carried out using vcf2maf v1.6.20, and the ENCODE blacklist was used to filter blacklist genes. Deleterious variants were selected for further analysis based on the annotation [19].

B. Copy number variant calling, Mutational signature analysis, Visualization, Oncogenic pathway, and Microsatellite instability analysis

Copy number variants were analyzed by CNVkit with batch option [20]. By using the CNVkit filter cn option to merge nearby values with the same called value, the results were changed. Genes with a copy number greater than 4 were classified as amplifying genes, whereas genes with 0 copies were classified as being deleted. Based on the UCSC reference, gene annotations were made.

Maftools' decomposition of the nucleotide substitution matrix allowed for the extraction of mutational signatures [21]. Elbow method was used to decide optimal number of signatures. Extracted signatures were compared to built-in COSMIC database. Further signature analysis was performed by SigProfilerExtractor and SigProfilerAssignment [22,23]. Visualization including ‘Oncoprint’, ‘Heatmap’ and ‘Lollipop plots’ were drawn by R package ComplexHeatmap and maftools. Additional graphs were visualized using R package ggplot2, ggsignif, ggradar, ggsci, gridtext, and gridExtra [24-26]. The oncogenic pathway was determined based on findings from

previous studies [27,28]. Microsatellite instability (MSI) is confirmed using MSIsensor, with a threshold of 3.5 distinguishing between microsatellite instability (MSI) and microsatellite stable (MSS) statuses [29,30].

C. Nanostring assay and analysis

The nCounter Analysis System (Nanostring Technologies, Seattle, WA) was used to screen for the expression of 55 immune-related genes. Counts were filtered by using negative probe using the nSolver software ver. 4.0 to remove outliers. Geometric mean of positive probe and housekeeping genes were used for normalizing data. Volcano plots were drawn by the ggplot2 R package.

The annotation for each profile was provided by Nanostring company. T-cell function genes include *CD2, CD27, CD274, CD38, CD3E, CD3G, CD80, CD86, CD8A, CTLA4, CXCL10, CXCL9, CXCR5, IDO1, IFNG, IL18, IRF1, LAG3, LCK,* and *TIGIT*. Chemokine genes include *CCL5, CX3CR1, CXCL10, CXCL13, CXCL9, CXCR5, CXCR6, IL2RG, IRF1,* and *STAT1*.

D. Statistical method

Statistical analysis was performed on R (ver 4.0.2). In order to analyze categorical variables, Fisher's exact test or Chi-square test will be used, and continuous variable was analyzed by t-test, Wilcoxon rank sum test or Kruskal Wallis test. Multiple testing correction was used by p.adjust function with Bonferroni correction.

E. Survival analysis

Survival analysis was conducted using the "survival" R package, employing the Kaplan-Meier method, log-rank test, and cox regression analysis. To ensure reliability, only data from the umbrella trial were utilized for the survival analysis. For additional PI3-kinase mutation analysis, only patients with PIK3CA SNV/indel or amplification were selected for the analysis.

4. Ethical statement

This study was conducted with the approval of the Institutional Review Board of 19 institutions in Korea. All patients provided written informed consent for genomic testing used in this study.

III. RESULTS

1. Patients′ clinical characteristics

Genomic and clinical information of 419 patients from the TRIUMPH trial was used in this study after prescreening. The baseline characteristic of patients is summarized on **Table 1**. Among the 419 patients, 276 (70%) had a history of smoking. The median age was 61 years, and most patients (88%, 51/419) were male. The primary tumor sites included oral cavity (n=145, 35%), hypopharynx (n=82, 20%), oropharynx (n=78, 19%) and larynx (n=75, 18%). HPV status was tested for 175 patients, 62 of whom (29%) were positive. The majority of the HPV-positive HNSCC was oropharyngeal cancer (43/62, 69%).

Table 1. Baseline characteristics of 419 patients prescreened in the TRIUMPH trial.

Variable	level	Overall	Oral cavity	Hypopharynx	Oropharynx	Larynx	Maxillary sinus	Nasal cavity
No.		419	145	82	78	75	23	16
Age		60.9 (11.3)	57.7 (13.5)	64.3 (9.7)	60.5 (8.3)	64.9 (8.1)	58.7 (10.3)	57.8 (13.7)
Gender	F	64 (15.5)	46 (31.9)	5 (6.2)	5 (6.6)	1 (1.4)	4 (17.4)	3 (18.8)
	M	350 (84.5)	98 (68.1)	76 (93.8)	71 (93.4)	73 (98.6)	19 (82.6)	13 (81.2)
Stage	1	28 (7.8)	13 (10.1)	1 (1.3)	5 (7.4)	7 (12.7)	0 (0.0)	2 (18.2)
	2	23 (6.4)	9 (7.0)	6 (8.0)	4 (5.9)	3 (5.5)	0 (0.0)	1 (9.1)
	3	53 (14.8)	20 (15.5)	8 (10.7)	14 (20.6)	5 (9.1)	3 (14.3)	3 (27.3)
	4A	222 (61.8)	77 (59.7)	48 (64.0)	38 (55.9)	37 (67.3)	18 (85.7)	4 (36.4)
	4B	12 (3.3)	5 (3.9)	6 (8.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (9.1)
	4C	21 (5.8)	5 (3.9)	6 (8.0)	7 (10.3)	3 (5.5)	0 (0.0)	0 (0.0)
Smoking status	Current smoker	66 (15.8)	23 (15.9)	20 (24.4)	12 (15.4)	8 (10.7)	1 (4.3)	2 (12.5)
	Former smoker	210 (50.1)	63 (43.4)	39 (47.6)	50 (64.1)	44 (58.7)	9 (39.1)	5 (31.2)
	Never smoker	119 (28.4)	51 (35.2)	20 (24.4)	13 (16.7)	17 (22.7)	11 (47.8)	7 (43.8)
HPV infection status	Positive	62 (14.8)	10 (6.9)	4 (4.9)	43 (55.1)	5 (6.7)	0 (0.0)	0 (0.0)
	Negative	150 (35.8)	58 (40.0)	39 (47.6)	16 (20.5)	25 (33.3)	7 (30.4)	5 (31.2)
	Not available	207 (49.4)	77 (53.1)	39 (47.6)	19 (24.4)	45 (60.0)	16 (69.6)	11 (68.8)

Numbers are No. (%) unless otherwise noted

2. Genomic and transcriptomic landscape of HNSCC

We analyzed the overall patterns of somatic mutations in HNSCC including somatic single nucleotide variants (SNVs), insertion/deletions (indels), and amplifications (**Fig. 2A**). The most frequently mutated genes were consistent with previous studies [14,31-33], including *TP53* (71%, 296/419), *CDKN2A* (27%, 112/419), *PIK3CA* (26%, 110/419), *FAT1* (23%, 95/419), and *EGFR* (16%, 69/419). We also noted recurrent mutations on hotspot sites, such as gain-of-function mutations in *PIK3CA* (p.E545K/A/G, p.E542K, and p.H1047R/L), and truncating mutations in *CDKN2A* (p.R80*, p.W110*, and p.X51_splice) (**Fig. 3**) [34].

Somatic copy number variations (CNVs) were frequently found in *CDKN2A* (10%), *PIK3CA* (15%), *EGFR* (13%), *FGFR1* (6%), *ATR* (6%), and *CCND1* (25%) (**Fig. 8**). In *CDKN2A*, almost CNVs were deletions, confirming its tumor suppressive role along with the truncating mutations. Other genes were mostly amplified. Transcriptomic analysis further revealed the effect of CNVs on the gene expression. We found clear correlation between the copy number gain and the increase expression in *EGFR* (p=2.20×10-16), *ERBB2* (p=2.63×10-9), and *PIK3CA* (p=0.0011) (**Fig. 4A**), except FGFR1. Overall, the increase was highest in *EGFR* (**Fig. 4B**). Indeed, *EGFR* amplification was associated with reduced survival compared to *PIK3CA* amplification (**Fig. 4C and 4D**). These results indicate that the use of copy number changes as a marker for targeted therapy might need further recalibrated by their impact on gene expression [35].

Mutational signature analysis identified four major signatures from the somatic mutations of HNSCC (**Fig. 1B and 1C**; see Methods). We found that four signatures matched to known COSMIC signatures with prosed etiology: SBS1 (spontaneous or enzymatic deamination of 5-cytosine), SBS2 (APOBEC Cytidine Deaminase), SBS15 (Defective DNA mismatch repair) and SBS31 (Prior chemotherapy treatment with platinum drugs). Notably, SBS15 has been known to be a surrogate marker for mismatch repair deficiency, activating cytotoxic T cells and alerting the immune

system [36,37]. Likewise, SBS31 has been associated with chemotherapy resistance [38]. Along with genetic aberrations, presentation of such signatures can be a marker for treatment, and a guide to personalized treatment approaches in HNSCC.

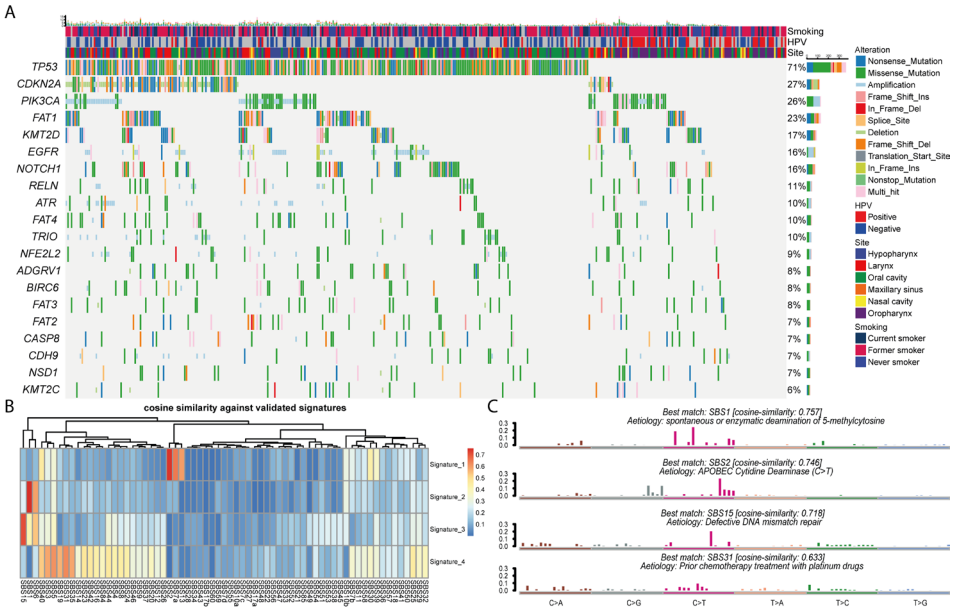

Figure 2. Overall genomic landscape and mutational signature of HNSCC.
(A) Overall mutational pattern thorough oncoprint with top 20 mutated genes. (B) Mutational signature analysis using De novo extraction of somatic mutations by non-negative matrix factorization. Extracted signatures were compared on full table of cosine similarities against COSMIC signatures. (C) The best match signatures were shown on plot.

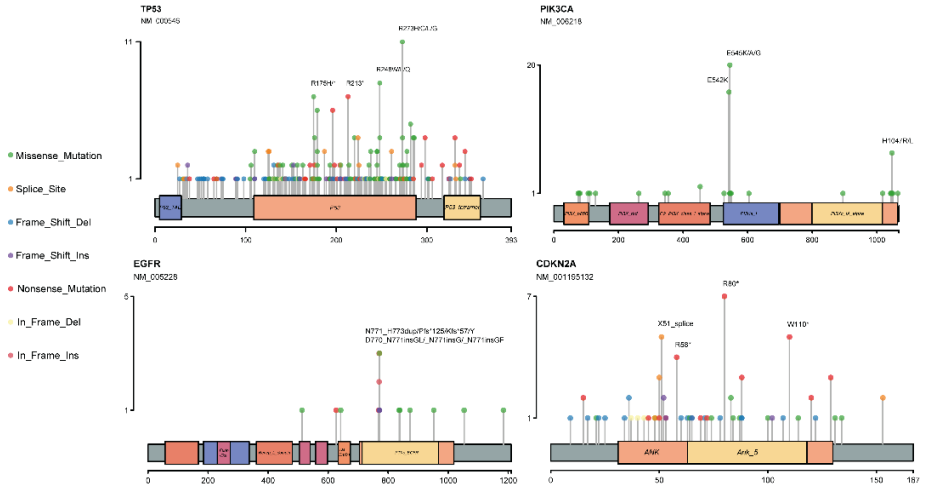

Figure 4. Lollipop plots of representative mutated genes in our cohort for amino acids change.

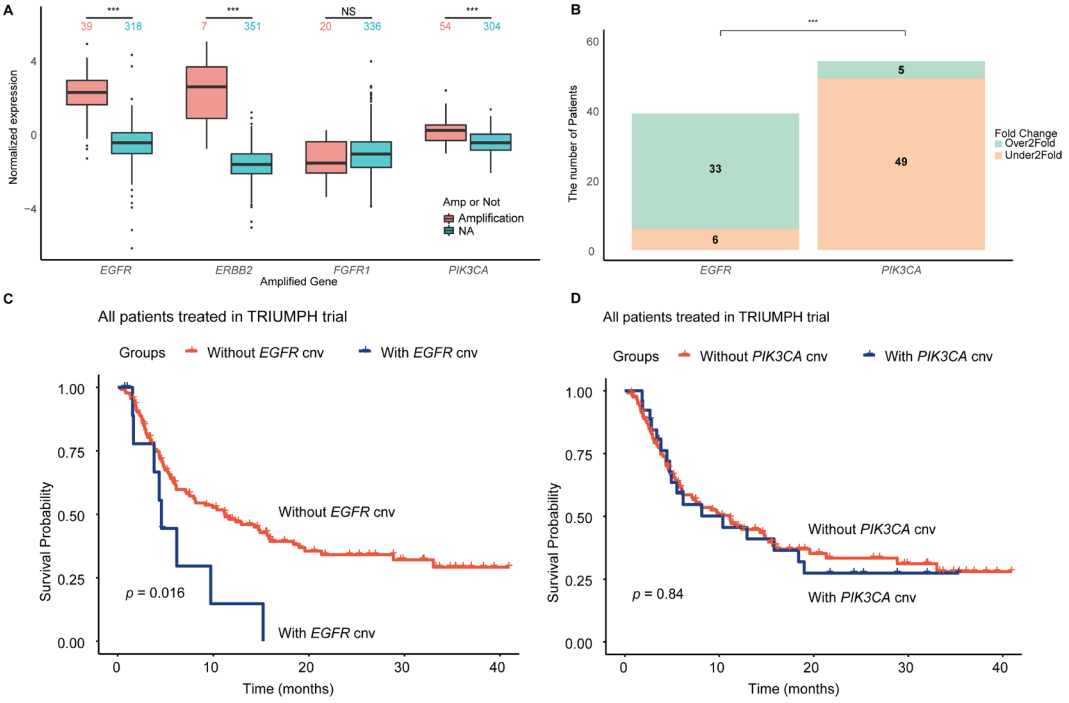

Figure 3. Gene amplification, transcriptomic patterns, and relevant clinical characteristics.
(A) The normalized expression difference according to gene amplification status. (B) The number of patients who show 2-fold increases expression compared to normal. Kaplan-Meier survival curves according to EGFR amplification (C) and PIK3CA amplification (D).

3. Genetic association with clinical outcome

Associations between genomic features and the clinical outcome were investigated from the treatment and responses[27]. In total of 179 patients (out of 419) were assigned to four different treatment arms based on the exhibiting genetic mutations and the eligibility criteria: PI3KCA inhibitor arm (Alpelisib, n=43), EGFR/HER2 inhibitor arm (Poziotinib, n=17), FGFR inhibitor arm (Nintedanib, n=10), and CDK4/6 inhibitor arm (Abemaciclib, n=33) (**Fig. 5** and **Fig. 1**). The other patients without eminent mutations were assigned to the immunotherapy arm (Duvalumab +/− Tremelimumab, n=76). Statistical analysis without group-level stratification identified TP53 and NOTCH1 as general indicators for a poor prognosis (multivariate Cox regression p=0.0046 and 0.0072 in *TP53* and *NOTCH1*, respectively), but not *CDKN2A* (p=0.27) (**Table 2**).

Most of the patients in the total cohort had mutations or copy number changes in the PI3K signaling pathway genes (*PIK3CA*, *PIK3CB*, *PIK3C2A*, *PIK3C3*, *PIK3CD*, or *PIK3R1*) (34.8%, n=146/419). Among them, *PIK3CA* is the mostly frequently mutated (75%, n=110/146) in three hotspot sites: E545K/A/G (20.0%, 22/110), E542K (14.5%, 16/110), and H1047R/L (6.4%, 7/110). The position of which had no difference on the clinical outcome in PIK3CA inhibitor arm (Fisher's exact test, p=1, **Fig. 5B**). Likewise, mutation type (SNV/indel vs amplification) (**Fig. 5C**) or presence of concurrent mutations (*PIK3CA* only vs. *PIK3CA* + accompanying mutations) (Fig. 2D) did not show associations of statistical significance. Conversely, mutations in the AKT-mTOR pathway were associated with improved prognosis (**Fig. 5E**). Moreover, we found that *NOTCH1* mutation (p=0.0037, log-rank test) (**Fig. 5F**) and *MYC* amplification (p=0.0016, log-rank test) (**Fig. 5G**) were associated with a poor prognosis. These findings suggest that potential genetic markers for patient survival in response to PIK3CA inhibitors, may reside outside of the PI3K signaling pathway.

In the EGFR/HER2 pathway, mutations in *EGFR* (58%), *ERBB2* (11%), *ERBB3* (6%) and *ERBB4* (18%) were most frequently observed (**Fig. 5H**). However,

no specific genetic factors were found to be associated with clinical outcome, including the presence of *EGFR* mutation in EGFR/HER2 inhibitor arm (amplification and SNV/indel) (p=0.15, **Fig. 5I**). Likewise, none of the genetic mutations within FGFR pathway, including *FGFR1* (53%), *FGFR2* (14%), *FGFR2* (16%), and *FGFR4* (4%) showed association with patient survival in FGFR inhibitor arm (**Fig. 5J** and **5K**). These findings may be attributed to the limited number of cases analyzed.

In the cell cycle pathway, *CDKN2A* (60%) and *CCDN1* (55%) were the mostly frequently mutated (**Fig. 5L**). Among them, we found a notable association pattern of *CDKN2A* mutation with patient survival, which leads to opposite prognosis depending on the mutation types; *CDKN2A* deletions were associated with a favorable prognosis (p=0.013, log-rank test), whereas SNVs/indels were linked to a poor prognosis in patients treated with CDK4/6 inhibitor (p=0.049, log-rank test) (**Fig. 5M** and **5N**). On the other hand, *CCND1* amplification did not show a significant relationship with survival (**Fig. 5O**). Furthermore, mutations in other signaling pathways, including PI3K, EGFR, and FGFR pathway, were not associated with patient prognosis (**Fig. 5P**, **5Q**, and **5R**). These findings highlight the complexities in mutation impacts on prognosis, emphasizing the importance of assessing not only the presence but also the type of mutations in developing effective prognostic markers.

Table 2. Hazard ratio (HR) for overall survival based on genetic alteration by Cox regression analysis of patients treated in TRIUMPH trial.

Presence of genetic alteration		N (%)	Univariate HR (Confidence Intervals)	Multivariate HR (Confidence Intervals)
TP53	No	63 (35.2)	-	-
	Yes	116 (64.8)	2.19 (1.38-3.46), p=0.001	2.02 (1.24-3.28), p=0.005
NOTCH1	No	156 (87.2)	-	-
	Yes	23 (12.8)	2.29 (1.23-4.29), p=0.009	2.38 (1.26-4.48), p=0.007
CDKN2A	No	133 (74.3)	-	-
	Yes	46 (25.7)	1.71 (1.06-2.76), p=0.027	1.33 (0.80-2.20), p=0.271

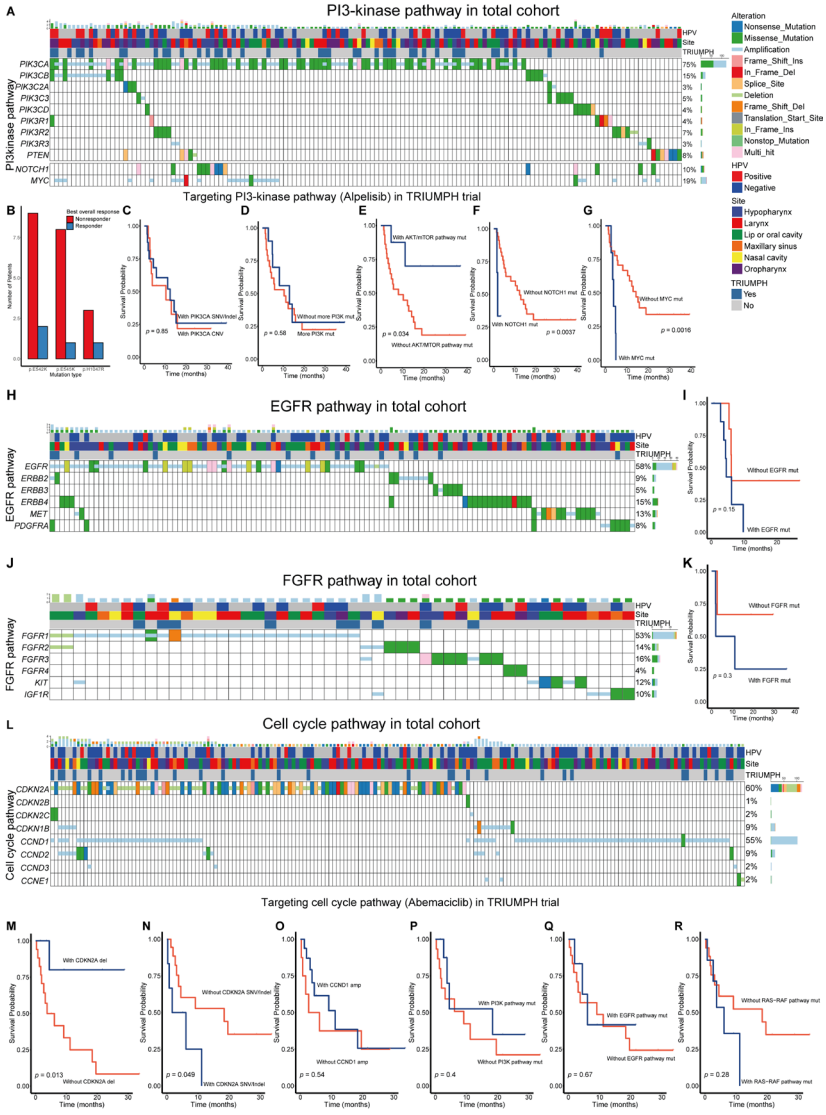

Figure 5 Potential therapeutic targeted pathways and genetic association with clinical outcome.
The mutational patterns of potential therapeutic targeted pathways include (A) the PI3K pathway, (H) the EGFR pathway, (J) the FGFR pathway, and (L) the Cell Cycle pathway. (B) The best overall response of Alpelisib based on *PIK3CA* mutation types. (C) Kaplan-Meier survival curves of patients with *PIK3CA* SNV/indel versus patients with *PIK3CA* amplification, in patients treated with Alpelisib. (D) Kaplan-Meier survival curves of patients with *PIK3CA* mutations with or without concurrent mutations on the PI3K pathway, in patients treated with Alpelisib. (E) Kaplan-Meier survival curves comparing patients with mutations in the AKT/mTOR pathway to those without, treated with Alpelisib. (F) Kaplan-Meier survival curves comparing patients with *NOTCH1* mutations to those without, treated with Alpelisib. (G) Kaplan-Meier survival curves comparing patients with *MYC* mutations to those without, treated with Alpelisib. (I) Kaplan-Meier survival curves comparing patients with *EGFR* mutations to those without, treated with Poziotinib. (K) Kaplan-Meier survival curves comparing patients with *FGFR* mutations to those without, treated with Nintedanib. (M) Kaplan-Meier survival curves comparing patients with *CDKN2A* deletions to patients without *CDKN2A* deletions, treated with Abemaciclib. (N) Kaplan-Meier survival curves comparing patients with *CDKN2A* SNV/indels to those without, treated with Abemaciclib. (O) Kaplan-Meier survival curves comparing patients with *CCND1* copy number alterations to those without, treated with Abemaciclib. (P) Kaplan-Meier survival curves comparing patients with mutations in the PI3K pathway to those without, treated with Abemaciclib. (Q) Kaplan-Meier survival curves comparing patients with mutations in the EGFR pathway to those without, treated with Abemaciclib. (R) Kaplan-Meier survival curves comparing patients with mutations in the RAS-RAF pathway to those without, treated with Abemaciclib.

4. Primary site associated genomic and transcriptomic characteristics

We conducted a deeper analysis on the genetic and transcriptomic features that exhibit primary site specificity. Site-level genomic landscape showed lower *TP53* mutation rate in oropharyngeal cancer ($p=1.61\times10^{-5}$, **Fig. 6A** and **6C**), which is believed to be influenced by higher HPV infection rate [39]. Also, *KMT2D, CASP8* and *FBXW7* show different frequencies according to primary sites. (p-value < 0.05 on Fisher exact test with Bonferroni correction on 30 top genes). In our cohort, the negative association between *TP53* mutation and HPV infection was independently confirmed (**Fig. 8A**). We also observed different CNV patterns among primary sites (**Fig. 7A** and **Fig. 7B**). Amplification of key driver genes was most prevalently observed in hypopharynx cancer (mean 26.8%) (**Fig. 7B**). In contrast, along with the *TP53* mutations, copy number variations showed different frequencies based on primary sites, including *CTTN, CCND1, SOX2, SOX2-OT, KLHL6, KDM6A, TP63, PIK3CA, PRKCI, PAK3, EGFR, EGFR-AS1, CDKN2A, ATR, EPH4, and SLC12A9* (**Fig. 6D**).

The differential frequencies of CNVs among primary sites were generally correlated with other variables, including HPV infection, *TP53* mutation status, smoking experience, and age (**Fig. 7C**), suggesting the potential factors for CNV burden. Higher CNV burden was shown in patients with *TP53* mutation (Fisher's exact $p = 5.938\times10^{-7}$, **Fig. 7D**), current/former smoker ($p=0.0496$, **Fig. 7E**), and age (Univariate logistic regression $p=5.44\times10^{-5}$, **Table 3**). In contrast, no significance was observed in HPV infection (Fisher's exact test $p=0.310$, **Fig. 8B**). Unlike previous reports [9], there was no correlation between smoking status and *TP53* mutation (**Fig. 8C**), implying that *TP53* mutation is an independent factor for higher CNVs. Indeed, *TP53* mutation and older age are the only significant factors in both Univariate (**Table 3**) and Multivariate logistic regression (**Table 4**), which explains the differential CNV frequencies among primary sites.

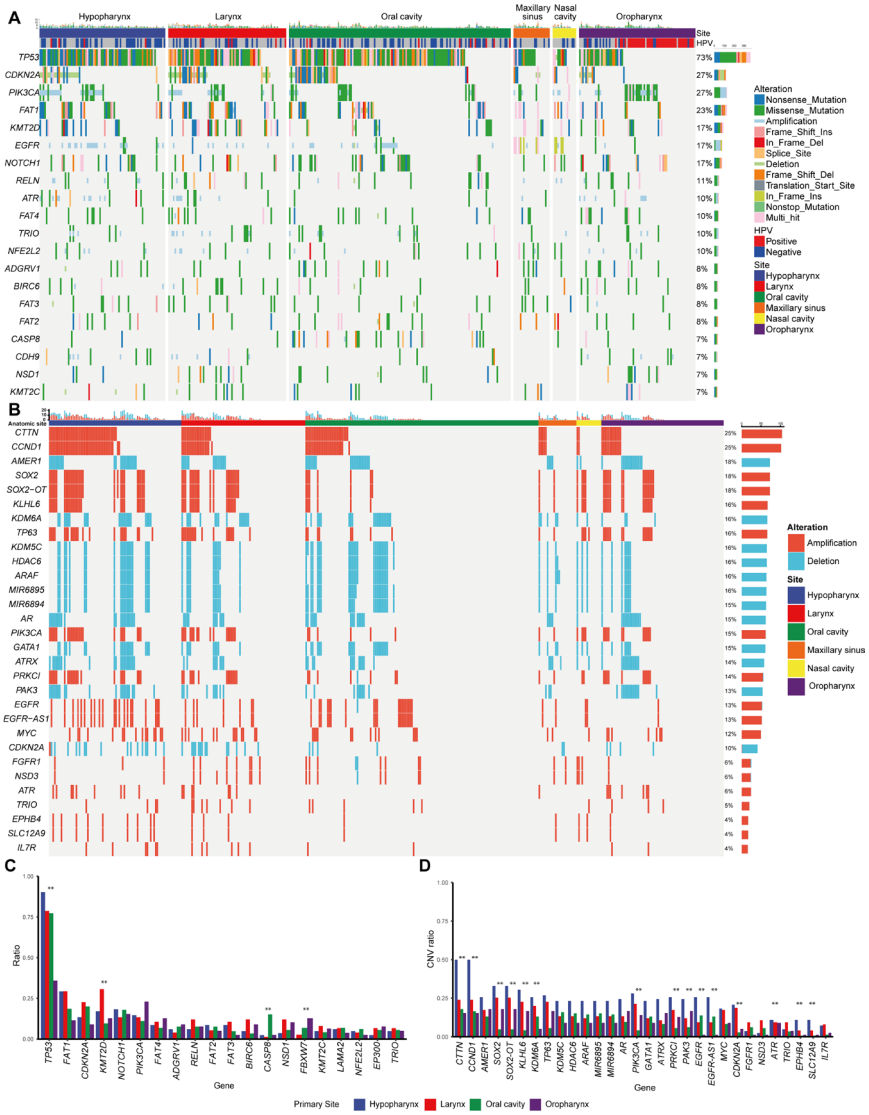

Figure 6. The mutational pattern according to primary sites.
(A) Somatic mutations only (B) Copy number variants only (C) The ratio of somatic mutation status according to primary sites, (D) the ratio of copy number variant status according to primary sites.

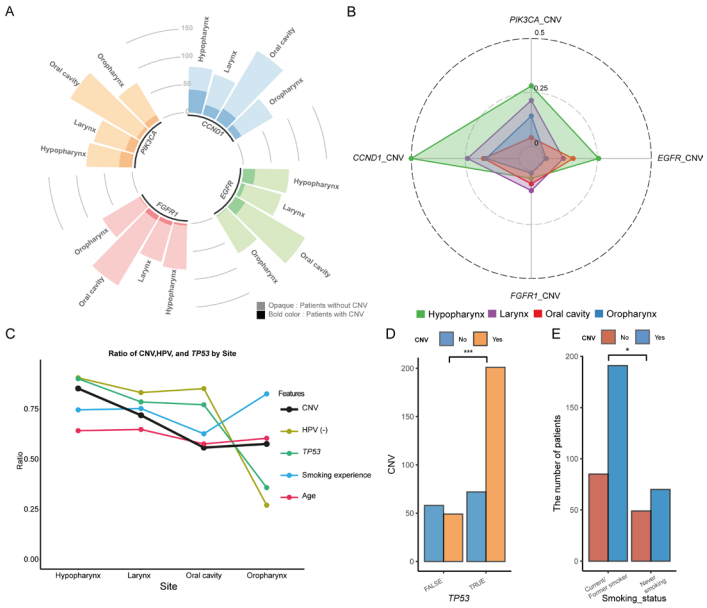

Figure 7. The determinants of copy number variation in HNSCC.
(A) The number of patients with CNV according to primary sites (B) The ratio of patients with CNV according to primary sites (C) The ratio of specific factors according to primary sites (D) The number of patients with CNV according to TP53 mutation status (E) The ratio of smoking status depending on the primary sites (F) The number of patients with CNV according to the smoking status. (** Fisher exact test with Bonferroni correction p-value < 0.05)

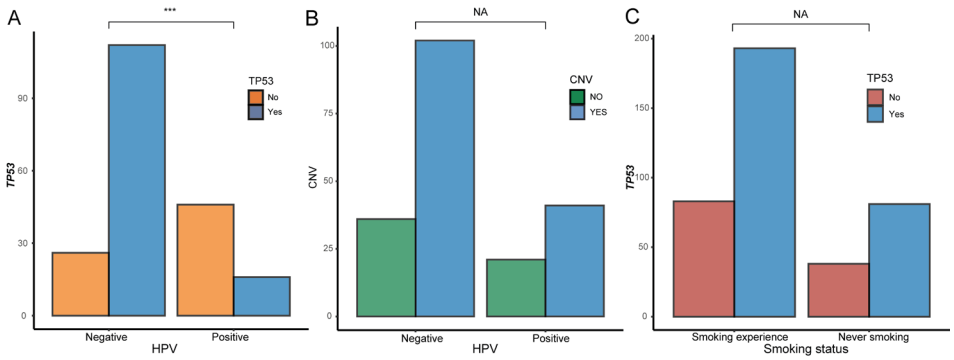

Figure 8. The factors affected CNV status in HNSCC patients.
(A)The number of patients with TP53 mutation according to HPV infection status (B)The number of patients with CNV according to smoking status (C)The number of patients with TP53 mutation according to smoking status

Table 3. Univariable logistic regression of CNV development

Characteristic	Group	N	OR[1]	95% CI[2]	p-value
Age		409	1.04	1.02, 1.06	<0.001
Smoking status	Never smoking	119	-	-	
	Current/Former smoker	276	1.57	1.01, 2.45	0.046
TP53	No	124	-	-	
	Yes	295	3.23	2.09, 5.03	<0.001

[1] OR = Odds Ratio, [2] CI = Confidence Interval

Table 4. Multivariable logistic regression of CNV development

Characteristic		N	OR[1]	95% CI[2]	p-value
Age		409	1.04	1.02, 1.06	<0.001
Smoking status	Never smoking	119	-	-	
	Current/Former smoker	276	1.50	0.92, 2.42	0.10
TP53	No	124	-	-	
	Yes	295	3.25	2.05, 5.21	<0.001

[1] OR = Odds Ratio, [2] CI = Confidence Interval

Notable site-specific transcriptomic features were mainly observed in genes involved in T-cell function activity (**Fig. 9A**) with *TP53* mutation further contributed to the decrease in anti-T cell activity (**Fig. 9B**). In consistent with previous studies, we found that *CCND1* amplification led to decreased T-cell function, even in *TP53* mutated patients [40,41]. In the clinical trial data, *TP53* mutation was associated with lower survival in immunotherapy treated patients (p=0.0064, log-rank test, **Fig. 9C**). These results suggest that *TP53* mutation and *CCND1* amplification may contribute to T-cell function suppression in HNSCC, highlighting the potential of combining cell cycle inhibitors with immunotherapy for HNSCC treatment.

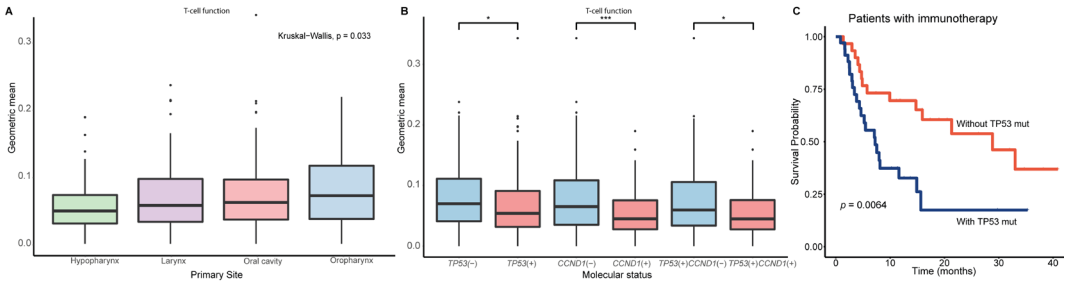

Figure 9. Immune profile of HNSCC.
(A) Geometric mean of T-cell function related cytokine according to primary sites. (B) Geometric mean of T-cell function related cytokines according to molecular status. (C) Kaplan-Meier progression free survival curves according to TP53 mutation status in patients who assigned to immunotherapy.

5. Subgroup analysis on oropharyngeal cancer

HPV infection has been known as the primary factor in the classification in oropharyngeal cancer [42]. Here, we analyzed genetic and clinical characteristics between HPV positive and negative subgroups of oropharyngeal cancer. As previously reported, *TP53* mutation rate was higher in HPV negative patients (11/16, 68.8%) than in HPV positive (4/43, 9.3%) (**Fig. 10A**). Notably, we observed mutual exclusiveness between *TP53* and *PIK3CA* SNV/indel in HPV negative patients (Fisher exact test, p=0.0018, **Fig. 10B**), confirming of the role of *PIK3CA* as driver in *TP53*-negative, HPV-negative oropharyngeal cancer [43,44]. In contrast, no such patterns were observed in HPV positive patients (p=0.43, **Fig. 10C**).

Transcriptomic patterns were also associated with HPV infection status (**Fig. 10D**). We found upregulation of genes involved in immune process and chemokine in HPV positive patients (Fig. 4D and 4E), including *CEACAM1, IL2RG, BTN1A2, CCL5, CXCL10, CXCR5,* and *CXCR6*. Likewise, interferon-gamma (*IFNG*) was up-regulated in HPV positive patients (**Fig. 11A**), as previously reported [45-47]. We found increase expression of T-cell function genes and cytotoxic cytokines in HPV infected patients (**Fig. 11B-D**), suggestive of favorable response to immunotherapy. Indeed, HPV positive patients who were initially treated with immunotherapy in arm 5 showed

prolonged survival outcomes compared with other targeted therapy in umbrella trial (**Fig. 10F**, log-rank test, p-value =0.064).

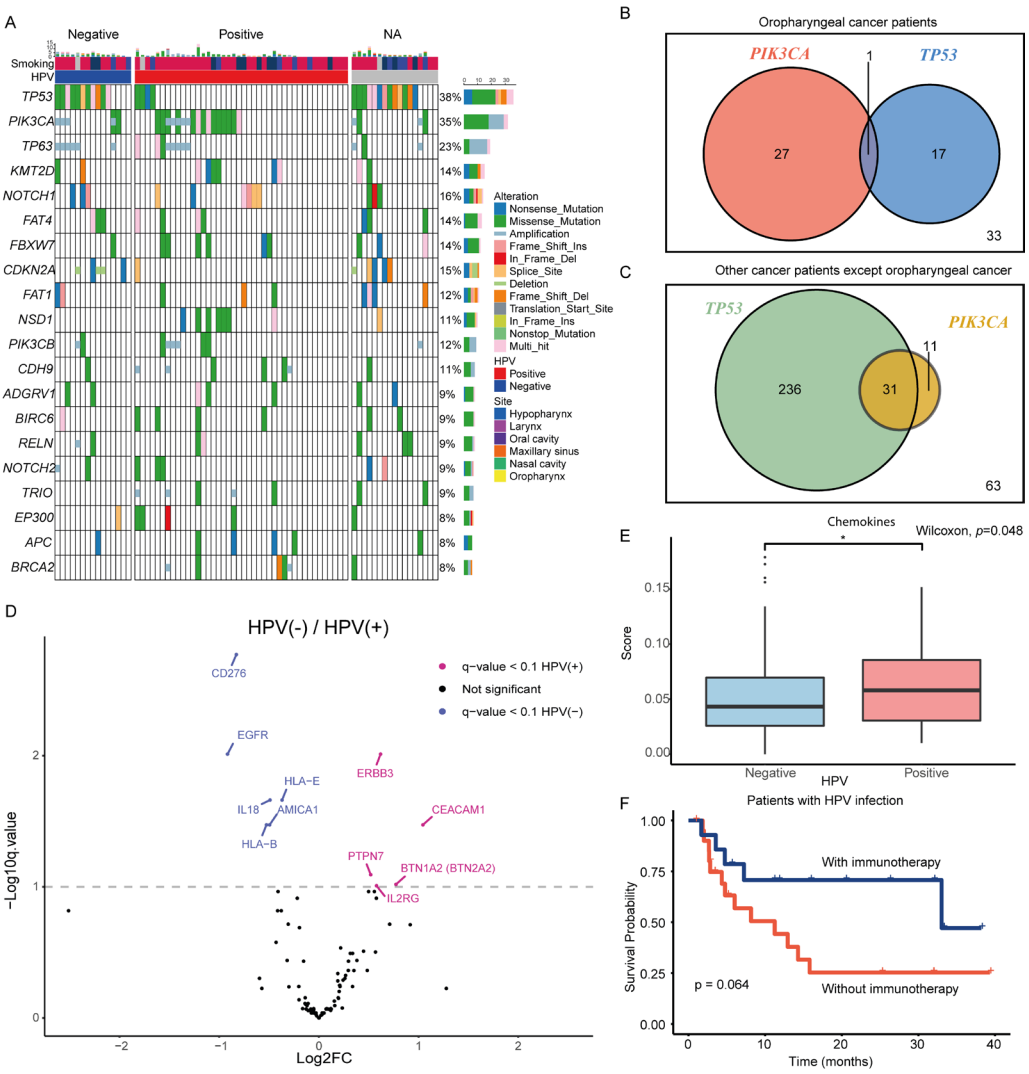

Figure 10. The characteristic of HPV-related HNSCC.
(A) Comparison of mutational pattern between HPV-positive and HPV-negative patients. (B) The relationship between SNV/indel of PIK3CA and TP53 in oropharyngeal HNSCC (C) The relationship between SNV/indel of PIK3CA and TP53 in other cancer patients. (D) The volcano plot representation of differential expression analysis between HPV positive and negative HNSCC patients. (E) The geometric mean of chemokine expression between HPV positive and negative HNSCC patients. (F) Kaplan-Meier survival curves of HPV positive and HPV negative HNSCC in patients who were assigned immunotherapy at first.

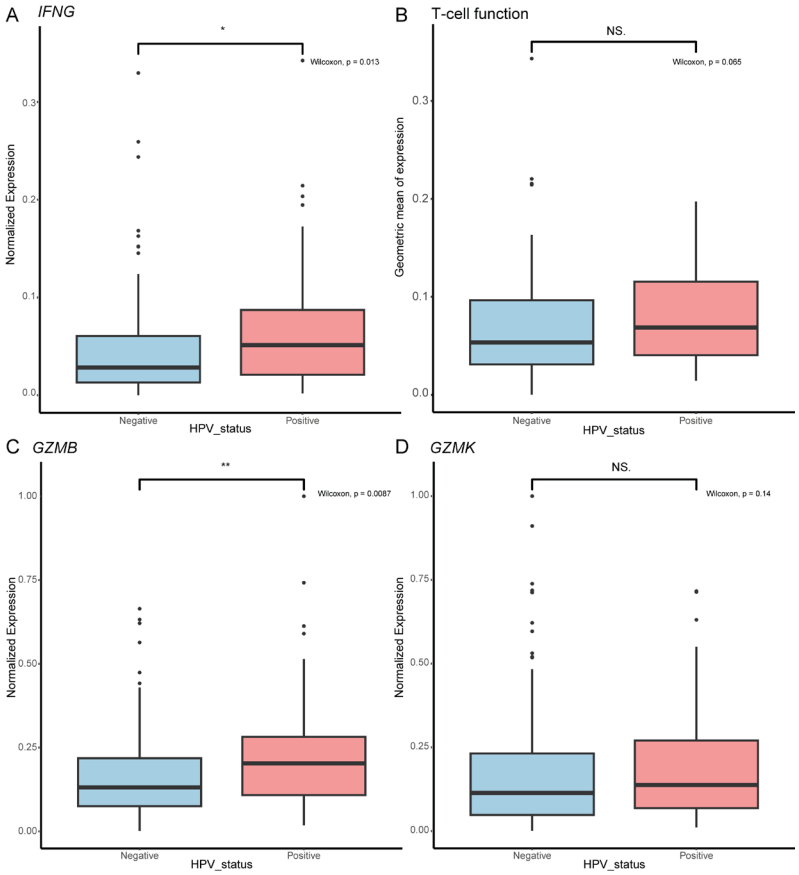

Figure 11. Immune profile of HPV-related HNSCC.
(A) The normalized expression of interferon-gamma (IFNG) according to HPV infection status. (B) Geometric mean of T-cell function related cytokines according to HPV infection status (C) The normalized expression of Granzyme B (GZMB) according to HPV infection status. (D) The normalized expression of Granzyme K (GZMK) according to HPV infection status

6. Revisiting association between smoking and oral cavity cancer

Smoking is a well-established carcinogenic factors of oral squamous cell carcinoma50. However, our cohort displayed the highest ratio of nonsmokers among patients with oral cavity cancer (**Fig. 12A**). To investigate the mutational signature associated with smoking, we analyzed a dataset comprising 90 patients who underwent both DNA and RNA sequencing as part of The Cancer Genome Atlas Program (TCGA). We assigned combined signatures to our cohort and TCGA data. Notably, patients with oral cavity cancer demonstrated a statistically significant lower prevalence of signature SBS4 compared to laryngeal cancer patients in the TRIUMPH cohort (Fisher exact test, p=0.042, **Fig. 12B**). Moreover, oral cavity cancer patients in the TCGA cohort did not exhibit the smoking-related mutational signature (p=0.00044, **Fig. 12C**). While this finding does not suggest that tobacco is unconnected to the development of oral cavity cancer, the evidence for smoking-induced mutations appears relatively less pronounced than in laryngeal cancer. Additionally, these results could indicate possible interactions between smoking and other factors, such as viral infections, in influencing cancer development.

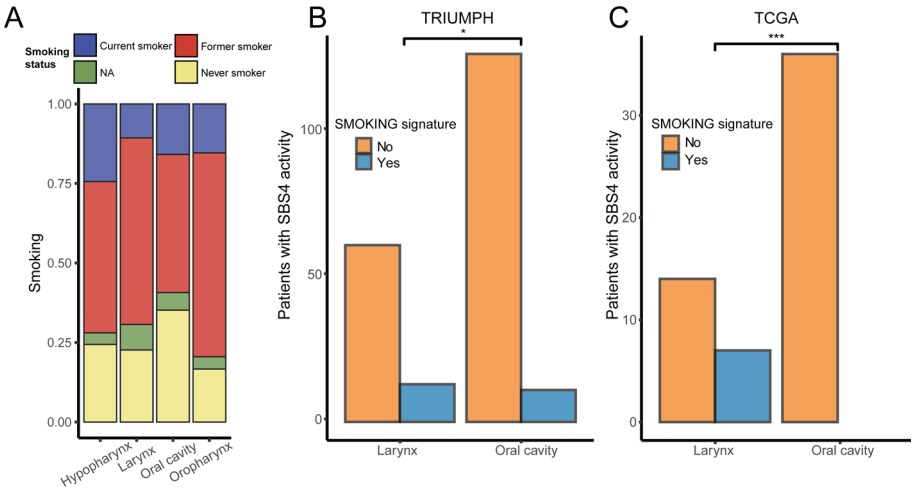

Figure 12. The various smoking features in oral cavity cancer.
(A) Ratio of smoking status depending on the primary sites. (B) The number of patients with or without the smoking-related signature (SBS4) comparing laryngeal cancer and oral cavity cancer in total cohort of TRIUMPH study. (C) The number of patients with or without the smoking-related signature (SBS4) comparing laryngeal cancer and oral cavity cancer in TCGA.

7. Genomic characteristics of young HNSCC patients

The etiology of young oral cavity cancer remains unclear51. We investigated the clinical and genomic characteristics of young (age ≤ 40 years) HNSCC patients who participated in the trial. Of the 21 patients studied, 16 (76%) were diagnosed with oral cavity cancer. The prevalence of *TP53* mutation in young patients (11/16, 69%) was comparable to that of in the older oral cavity cancer group (101/129, 78%) (Fisher exact test, p=0.28, **Fig. 13A** and **13C**). Additionally, there was no significant difference in the smoking history between the young (6/14, 43%) and older (78/121, 64%) groups; in fact, the younger group included a higher proportion of nonsmokers than the older group (Fisher exact test, p=0.15, **Fig. 13D**). None of the patients exhibited notable recurrent germline variants (**Fig. 13B**), denying the presence of congenital genetic factors in susceptibility. Also, all patients who underwent HPV testing were negative (6 out of **6, Fig. 13E**). Although the mismatch repair (MMR) signature (SBS15 and SBS44) was observed in the signature analysis, no significant differences were found in microsatellite instability (MSI) between the age groups (p=0.52, **Fig. 13F**). This comprehensive analysis suggests that the higher prevalence of oral cavity cancer in younger patients hint at heterogeneous origins beyond known genetic and other risk factors, warranting further exploration.

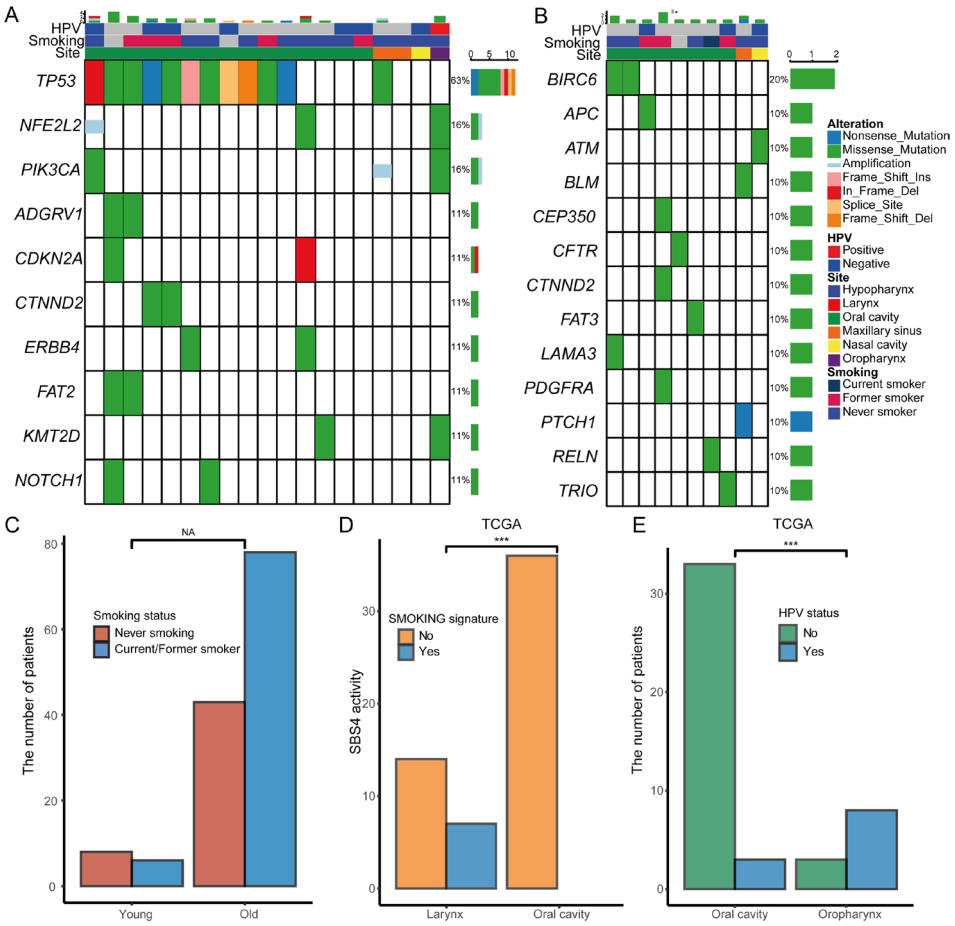

Figure 13. HNSCC patients with young age and Oral squamous cell carcinoma.
(A) Somatic mutations of patients under 40 years of age. (B) Germline variants of patients under 40 years of age. (C) The number of patients according TP53 mutation status between young and old patients. (D) The number of patients according smoking status between young and old patients. (E) The number of patients according HPV infection status between young and old patients. (F) The number of patients according microsatellite instability status (MSS versus MSI) between young and old patients.

IV. DISCUSSION

The genomic alteration of overall study included *TP53, CDKN2A, PIK3CA, FAT1,* and *EGFR* which are like previous studies [9,14]. We focused on the difference in genomic landscape according to primary cancer sites and many somatic mutations including *TP53, FAT1, KMT2D, CASP8,* and *FBXW7* were significantly enriched depending on the primary sites. The pattern of copy number variations was notably different based on the primary site of cancer and exhibited a more significant contrast. The primary factor in our cohort was *TP53* mutation status which was affected by HPV infection status. If a tumor has *TP53* mutation, it is more likely to have copy number variants possibly due to genomic instability. [48,49]

Smoking status was another determinant factor of copy number variants in our cohort. In TCGA data, *TP53* mutation, *CDKN2A* loss of function, and chromosome 3q amplification were frequently co-occurring with heavy smoking patients [9]. Also, previous report showed that cigarette smoking increased copy number alterations in non-small cell lung cancer patients [50]. In our data, the high proportion of non-smokers in lip or oral cavity cancer may explain the low copy number variants. We thought that age factor reflected the time from the occurrence of a key driver mutation, such as *TP53* mutation, to the accumulation of copy number variants. Because there was a relatively high proportion of young patients in the lip or oral cavity cancer group, they showed low copy number variants compared to hypopharyngeal or laryngeal cancer patients. To determine the variables influencing patient outcomes, survival data and molecular status were examined. We discovered a connection between *EGFR* amplification and elevated expression that was connected to worse patient survival. Remarkably, this link persisted even when only patients with *TP53* mutations were taken into account, albeit precise evidence for this discovery was not presented. The strong association between *EGFR* amplification and expression suggests a potential link to the efficacy of cetuximab in recurrent or metastatic HNSCC [6].

We have established a connection between the genomic landscape and the clinical

outcomes of targeted therapy in umbrella trial. Our findings reveal numerous significant elements, including previously unidentified abnormalities in key signaling networks and specific mutations that are associated with a poor response to targeted therapy. Notably, we discovered that mutations in the AKT-mTOR signaling pathway correlated with a favorable prognosis in treatments targeting the *PIK3CA* gene. In contrast, *NOTCH1* mutations and *MYC* amplifications were indicative of poor prognosis in PIK3CA-targeted therapies. This is particularly interesting as mutations in the AKT/mTOR pathway have previously been linked to improved survival with immune checkpoint inhibitors [51], suggesting these mutations might influence anti-tumor immunity during treatment with PIK3CA inhibitors. Conversely, *NOTCH1* and *MYC* mutations are known to confer resistance to PI3K inhibitors [52,53]. Interestingly, *CDKN2A* mutations did not correlate with poor prognosis in the multivariable Cox proportional hazard regression analysis of the entire TRIUMPH trial. However, patients with *CDKN2A* SNV/indel mutations exhibited poor outcomes when treated with CDK4/6 inhibitors. Our study highlights that not only the targeted mutations but also those related to the targeted pathway, or even mutations outside the targeted pathway, may influence patient survival. Hence, comprehensive mutation profiling, extending beyond the target pathway, is essential for optimizing targeted therapy.

To determine the variables influencing patient outcomes, survival data and molecular status were examined. We discovered a connection between *EGFR* amplification and elevated expression that was connected to worse patient survival. Remarkably, this link persisted even when only patients with *TP53* mutations were taken into account, albeit precise evidence for this discovery was not presented. The strong association between *EGFR* amplification and expression suggests a potential link to the efficacy of cetuximab in recurrent or metastatic HNSCC [6].

We have established a connection between the genomic landscape and the clinical outcomes of targeted therapy in umbrella trial. Our findings reveal numerous significant elements, including previously unidentified abnormalities in key signaling networks and

specific mutations that are associated with a poor response to targeted therapy. Notably, we discovered that mutations in the AKT-mTOR signaling pathway correlated with a favorable prognosis in treatments targeting the *PIK3CA* gene. In contrast, *NOTCH1* mutations and *MYC* amplifications were indicative of poor prognosis in PIK3CA-targeted therapies. This is particularly interesting as mutations in the AKT/mTOR pathway have previously been linked to improved survival with immune checkpoint inhibitors[51], suggesting these mutations might influence anti-tumor immunity during treatment with PIK3CA inhibitors. Conversely, *NOTCH1* and *MYC* mutations are known to confer resistance to PI3K inhibitors [52,53]. Interestingly, *CDKN2A* mutations did not correlate with poor prognosis in the multivariable Cox proportional hazard regression analysis of the entire TRIUMPH trial. However, patients with *CDKN2A* SNV/indel mutations exhibited poor outcomes when treated with CDK4/6 inhibitors. Our study highlights that not only the targeted mutations but also those related to the targeted pathway, or even mutations outside the targeted pathway, may influence patient survival. Hence, comprehensive mutation profiling, extending beyond the target pathway, is essential for optimizing targeted therapy.

HPV infection is key factor of pathogenesis of HNSCC. As the presence of the *TP53* mutation depends on the HPV infection status, the genomic profile of HNSCC is also classified according to HPV infections status. Because HPV-positive HNSCC usually do not have *TP53* mutation, it is less heterogenous. Without *TP53* mutation, somatic mutation of *PIK3CA* seems to have a more important role in pathogenesis in HPV-positive HNSCC. All responders who were treated with immunotherapy without *TP53* mutation had a somatic mutation of *PIK3CA* supporting this idea. In our cohort, HPV-positive HNSCC showed higher expression of chemokine-related cytokine with upregulation of *IFNG* and cytotoxic cytokine genes. We observed that HPV-positive HNSCC had a good prognosis in patients treated with immunotherapy. In our study, we observed that *TP53* mutation had a negative impact on the efficacy of immunotherapy, and one possible explanation for this finding is the presence of *CCND1* amplification. *CCND1* amplification has been shown to suppress T-cell activity and may contribute to reduced progression-free survival in patients

undergoing immunotherapy [54].

In our examination of germline variants, we did not discover any distinct patterns or recurrent variants based on primary sites or within the subgroup of young patients. However, the prevalence of *TP53* mutations remains high among young patients with oral cavity cancer, suggesting that another factor may contribute to the development of this disease. Interestingly, these young patients did not exhibit a high prevalence of HPV infection, nor did they have a history of current or former smoking. Our analysis of a selected TCGA dataset revealed heterogeneity in the underlying causes of oral cavity cancer. Mutational signature analysis identified the presence of a DNA Mismatch Repair (MMR) signature and a smoking signature; however, these signatures were not enriched in oral cavity cancer patients. The diversity in the origin of oral cancer complicates our understanding of the pathogenesis of oral cavity cancer, particularly in young patients. Therefore, the quest for a more sophisticated methodology to elucidate the pathogenesis of oral cavity cancer in young patients is needed.

There were several limitations in our study. First, we performed target gene sequencing and expression data, and these may restrict our analysis. Especially, mutational signature analysis can be affected by this restriction. Second, the time point at which the results were obtained was at the initial, and cancer may have changed from the initial data. Third, survival analysis should be affected by treatment, making the analysis more complicated. The number of subgroups was relatively small, and some data do not reach statistical significance.

In conclusion, our study conducted a large-scale genomic analysis of patients with HNSCC. We identified several genetic traits that were associated with clinical features. Despite the challenges posed by the heterogeneity of HNSCC, our findings have provided valuable insights that can potentially guide the treatment of real-world patients based on detailed genomic profiling.

V. CONCLUSION

In conclusion, we performed large-scale genomic analysis of HNSCC patients. We identified several genomic characteristics according to clinical features. Although original clinical study had difficulty in analyzing results due to heterogeneity of HNSCC, it was confirmed that our results were reflected in the treatment of actual patients through detailed genomic decomposition.

REFERENCES

1. Patterson RH, Fischman VG, Wasserman I, Siu J, Shrime MG, Fagan JJ, et al. Global Burden of Head and Neck Cancer: Economic Consequences, Health, and the Role of Surgery. Otolaryngology–Head and Neck Surgery 2020;162:296-303.
2. Burtness B, Harrington KJ, Greil R, Soulières D, Tahara M, de Castro G, Jr., et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. Lancet 2019;394:1915-28.
3. Ferris RL, Blumenschein G, Jr., Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. N Engl J Med 2016;375:1856-67.
4. Cohen EEW, Soulières D, Le Tourneau C, Dinis J, Licitra L, Ahn MJ, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. Lancet 2019;393:156-67.
5. Mehra R, Seiwert TY, Gupta S, Weiss J, Gluck I, Eder JP, et al. Efficacy and safety

of pembrolizumab in recurrent/metastatic head and neck squamous cell carcinoma: pooled analyses after long-term follow-up in KEYNOTE-012. Br J Cancer 2018;119:153-9.

6. Vermorken JB, Mesia R, Rivera F, Remenar E, Kawecki A, Rottey S, et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med 2008;359:1116-27.
7. Hitt R, Irigoyen A, Cortes-Funes H, Grau JJ, García-Sáenz JA, Cruz-Hernandez JJ. Phase II study of the combination of cetuximab and weekly paclitaxel in the first-line treatment of patients with recurrent and/or metastatic squamous cell carcinoma of head and neck. Ann Oncol 2012;23:1016-22.
8. Cabezas-Camarero S, Cabrera-Martín MN, Merino-Menéndez S, Paz-Cabezas M, García-Barberán V, Sáiz-Pardo Sanz M, et al. Safety and Efficacy of Cetuximab-Based Salvage Chemotherapy After Checkpoint Inhibitors in Head and Neck Cancer. Oncologist 2021;26:e1018-e35.
9. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576-82.
10. Qin T, Li S, Henry LE, Liu S, Sartor MA. Molecular Tumor Subtypes of HPV-Positive Head and Neck Cancers: Biological Characteristics and Implications for Clinical Outcomes. Cancers (Basel) 2021;13.
11. Saada-Bouzid E, Peyrade F, Guigay J. Molecular genetics of head and neck squamous cell carcinoma. Curr Opin Oncol 2019;31:131-7.
12. Hedberg ML, Goh G, Chiosea SI, Bauman JE, Freilino ML, Zeng Y, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. Journal of Clinical Investigation 2015;126:169-80.
13. Ju Y, Wu X, Wang H, Li B, Long Q, Zhang D, et al. Genomic Landscape of Head and Neck Squamous Cell Carcinoma Across Different Anatomic Sites in Chinese Population. Front Genet 2021;12:680699.
14. Lim SM, Cho SH, Hwang IG, Choi JW, Chang H, Ahn MJ, et al. Investigating the

Feasibility of Targeted Next-Generation Sequencing to Guide the Treatment of Head and Neck Squamous Cell Carcinoma. Cancer Res Treat 2019;51:300-12.

15. Lee J, Kim ST, Kim K, Lee H, Kozarewa I, Mortimer PGS, et al. Tumor Genomic Profiling Guides Patients with Metastatic Gastric Cancer to Targeted Treatment: The VIKTORY Umbrella Trial. Cancer Discov 2019;9:1388-405.
16. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-i90.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 2010;20:1297-303.
18. Kandoth C. mskcc/vcf2maf: vcf2maf v1.6.19. 2020.
19. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific reports 2019;9:9354-.
20. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLOS Computational Biology 2016;12:e1004873.
21. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747-56.
22. Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genomics 2022;2:100179.
23. Díaz-Gay M, Vangara R, Barnes M, Wang X, Islam SMA, Vermes I, et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. bioRxiv 2023:2023.07.10.548264.
24. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016.
25. Ahlmann-Eltze C, Patil I. ggsignif: R Package for Displaying Significance

Brackets for 'ggplot2'. PsyArXiv; 2021.

26. Xiao N. ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'. 2021.
27. Keam B, Hong MH, Shin SH, Heo SG, Kim JE, Ahn HK, et al. Personalized Biomarker-Based Umbrella Trial for Patients With Recurrent or Metastatic Head and Neck Squamous Cell Carcinoma: KCSG HN 15-16 TRIUMPH Trial. Journal of Clinical Oncology 2023:JCO.22.02786.
28. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell 2018;173:321-37.e10.
29. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics 2014;30:1015-6.
30. Jia P, Yang X, Guo L, Liu B, Lin J, Liang H, et al. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. Genomics, Proteomics & Bioinformatics 2020;18:65-71.
31. Dubot C, Bernard V, Sablin MP, Vacher S, Chemlali W, Schnitzler A, et al. Comprehensive genomic profiling of head and neck squamous cell carcinoma reveals FGFR1 amplifications and tumour genomic alterations burden as prognostic biomarkers of survival. Eur J Cancer 2018;91:47-55.
32. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. Science 2011;333:1157-60.
33. Klinakis A, Rampias T. TP53 mutational landscape of metastatic head and neck cancer reveals patterns of mutation selection. EBioMedicine 2020;58:102905.
34. Martínez-Sáez O, Chic N, Pascual T, Adamo B, Vidal M, González-Farré B, et al. Frequency and spectrum of PIK3CA somatic mutations in breast cancer. Breast Cancer Research 2020;22:45.

35. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell 2020;182:200-25.e35.
36. Amodio V, Mauri G, Reilly NM, Sartore-Bianchi A, Siena S, Bardelli A, et al. Mechanisms of Immune Escape and Resistance to Checkpoint Inhibitor Therapies in Mismatch Repair Deficient Metastatic Colorectal Cancers. Cancers; 2021.
37. Ma J, Setton J, Lee NY, Riaz N, Powell SN. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nature Communications 2018;9:3292.
38. Brady SW, Gout AM, Zhang J. Therapeutic and prognostic insights from the analysis of cancer mutational signatures. Trends in Genetics 2022;38:194-208.
39. Zhou G, Liu Z, Myers JN. TP53 Mutations in Head and Neck Squamous Cell Carcinoma and Their Impact on Disease Progression and Treatment Response. J Cell Biochem 2016;117:2682-92.
40. Chen Y, Huang Y, Gao X, Li Y, Lin J, Chen L, et al. CCND1 Amplification Contributes to Immunosuppression and Is Associated With a Poor Prognosis to Immune Checkpoint Inhibitors in Solid Tumors. Frontiers in immunology 2020;11:1620-.
41. Liu J, Lin J, Wang X, Zheng X, Gao X, Huang Y, et al. CCND1 Amplification Profiling Identifies a Subtype of Melanoma Associated With Poor Survival and an Immunosuppressive Tumor Microenvironment. Frontiers in Immunology 2022;13.
42. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 2011;29:4294-301.
43. Chiosea SI, Grandis JR, Lui VWY, Diergaarde B, Maxwell JH, Ferris RL, et al. PIK3CA, HRAS and PTEN in human papillomavirus positive oropharyngeal squamous cell carcinoma. BMC Cancer 2013;13:602.
44. Nichols AC, Palma DA, Chow W, Tan S, Rajakumar C, Rizzo G, et al. High

Frequency of Activating PIK3CA Mutations in Human Papillomavirus–Positive Oropharyngeal Cancer. JAMA Otolaryngology–Head & Neck Surgery 2013;139:617-22.

45. Bhatia R, Kavanagh K, Stewart J, Moncur S, Serrano I, Cong D, et al. Host chemokine signature as a biomarker for the detection of pre-cancerous cervical lesions. Oncotarget 2018;9:18548-58.
46. Garrido F, Wild CM, Mittelberger J, Dobler F, Schneider M, Ansorge N, et al. The Role of Chemokines in Cervical Cancers. Medicina (Kaunas, Lithuania) 2021;57:1141.
47. Hu T, Yang P, Zhu H, Chen X, Xie X, Yang M, et al. Accumulation of invariant NKT cells with increased IFN-γ production in persistent high-risk HPV-infected high-grade cervical intraepithelial neoplasia. Diagnostic Pathology 2015;10:20.
48. Eischen CM. Genome Stability Requires p53. Cold Spring Harb Perspect Med 2016;6.
49. Hanel W, Moll UM. Links between mutant p53 and genomic instability. J Cell Biochem 2012;113:433-9.
50. Huang Y-T, Lin X, Liu Y, Chirieac LR, McGovern R, Wain J, et al. Cigarette smoking increases copy number alterations in nonsmall-cell lung cancer. Proceedings of the National Academy of Sciences 2011;108:16345-50.
51. Cheng L, Wang Y, Qiu L, Chang Y, Lu H, Liu C, et al. mTOR pathway gene mutations predict response to immune checkpoint inhibitors in multiple cancers. Journal of Translational Medicine 2022;20:247.
52. Dail M, Wong J, Lawrence J, O'Connor D, Nakitandwe J, Chen SC, et al. Loss of oncogenic Notch1 with resistance to a PI3K inhibitor in T-cell leukaemia. Nature 2014;513:512-6.
53. Bhin J, Yemelyanenko J, Chao X, Klarenbeek S, Opdam M, Malka Y, et al. MYC is a clinically significant driver of mTOR inhibitor resistance in breast cancer. Journal of Experimental Medicine 2023;220:e20211743.

54. Farah CS. Molecular landscape of head and neck cancer and implications for therapy. Ann Transl Med 2021;9:915.

ABSTRACT(IN KOREAN)

원발위치 및 임상 특성에 따른 두경부암의 게놈 형태와 이를 이용한 치료의 활용

<지도교수 김 상 우 >

연세대학교 대학원 의학과

황 신 원

두경부암은 구강내, 인두, 또는 후두에서 발생하는 암으로 흔하게 발생하는 암으로 알려져 있다. 두경부암을 대상으로 한 임상연구 (TRIUMPH)는 다기관, 2상, 우산형 임상연구로 본 연구팀 전향적으로 선별되는 과정을 거친 419명의 두경부암 환자의 유전체 및 임상적 특성을 이용한 분석을 진행하였다. 대상 환자의 유전체 분석을 통해 발견된 유전변이는 *TP53, CDKN2A, PIK3CA, FAT1, EGFR*등에 존재하였으며 이는 기존의 연구와 일치하였다. 임상연구 각 군 내에서는 *PI3KCA*를 억제하는 경우에선 AKT-mTOR 신호전달경로에 유전변이가 있는 경우 좋은 예후를 보였으나, *NOTCH1* 유전자의 유전변이나 *MYC* 유전자의 증폭 등이 나쁜 예후와 연관되어 있는 것을 확인하였다. 반대로 CDK4/6 억제제를 사용한 그룹에서는 *CDKN2A*의 유전변이 종류와 예후가 연관되어 있는 것도 확인할 수 있었다. 이를 통해 표적치료에 있어서, 우리가 표적으로 삼은 유전자의 변이뿐만 아니라, 같은 신호 전달경로에 있는 변이들, 그리고 더 나아가, 다른 신호전달경로의 변이들도 예후에 영향을 줄 수 있는 것을 확인할 수 있었다. 전반적인 게놈 형태가 특히 두경부암의 원발 위치에 크게 영향을 받는 것을 확인하였으며 특히 복제 수 변이는 *TP53* 유전변이 여부, 나이, 그리고 흡연 여부에 영향을 받는 것을 확인할 수 있었다. 부분집단 분석에서는 인간유두종 바이러스 관련 두경부암의 경우 구인두암에서

TP53 유전자와 *PIK3CA* 유전자의 단일염기서열변이 혹은 삽입/결실 변이가 서로 상호 배타적으로 존재함을 확인하였다. 또한 인간유두종 바이러스에 감염이 있는 경우 케모카인 및 IFNG 같은 유전자 발현의 증가 등이 관찰되며, 이런 결과가 반영되듯이, 인간유두종 바이러스를 가진 환자에서는 면역항암제 치료를 시행한 환자들이 예후가 좋은 것을 확인하였다. 구강암의 부분집단 분석에서는 먼저 기존에 알려져 있는 담배와의 연관성과는 달리, 구강암에서 상대적으로 다른 위치에서 발생한 두경부암에 비해 높은 비흡연자 비율을 보였다. 이와 연계되어 담배와 관련된 돌연변이 시그니처도 후두암 등과 비교하였을 때 두드러지지 않았다. 전체 코호트 중에 40세 이하의 젊은 환자에서 대부분을 차지하는 구강암의 경우에는 특별히 반복되는 생식세포 변이나 높은 흡연 비율 등이 없이 많은 환자에서 *TP53* 유전변이를 젊은 나이에도 가지고 있는 것을 확인할 수 있었지만 마찬가지로 젊은 환자에서 높은 흡연자 비율이나 인간유두종 바이러스 감염, 미소위성체 불안정성 등이 높은 비율로 발견되지 않아 좀더 다양한 원인으로부터 유래되는 것으로 보이며, 추가적인 연구가 필요할 것으로 사료된다. 이러한 게놈 프로파일링을 통하여 두경부암에서의 유전체의 역할에 대한 이해를 높이고 임상적 활용에 대한 통찰을 확인하였다.

핵심되는 말 : 두경부암, 게놈 형태

PUBLICATION LIST

1. Kim JH, Hwang S, Son H, Kim D, Kim IB, Kim M-H, et al. Analysis of low-level somatic mosaicism reveals stage and tissue-specific mutational features in human development. PLOS Genetics 2022;18:e1010404.
2. Keam B, Hong MH, Shin SH, Heo SG, Kim JE, Ahn HK, et al. Personalized Biomarker-Based Umbrella Trial for Patients With Recurrent or Metastatic Head and Neck Squamous Cell Carcinoma: KCSG HN 15-16 TRIUMPH Trial. Journal of Clinical Oncology 2023:JCO.22.02786.